Behavioral Interventions

WILEY

RESEARCH ARTICLE OPEN ACCESS

Teaching Mathematics to Autistic Students in Special Education by Integrating Learn Units Within Precision Teaching

Natalie Leow-Dyke[1] | Rhys Jones[1] | Maria Rossi[1] | Athanasios Vostanis[2]

[1]Jigsaw CABAS School, Cranleigh, Surrey, UK | [2]Tizard Centre, University of Kent, Canterbury, Kent, UK

Correspondence: Athanasios Vostanis (a.vostanis@kent.ac.uk)

Received: 23 November 2025 | **Revised:** 2 June 2026 | **Accepted:** 8 June 2026

Keywords: behavioral fluency | discrete trial teaching | learning disabilities | money skills | standard celeration chart

ABSTRACT

Autistic students with co-occurring conditions often need additional support in mathematics. Precision Teaching (PT) has shown promise in this area, but its application has primarily focused on typically developing students, employing strategies, such as choral responding, that are less accessible to those with complex needs. Evidence is limited on PT's effectiveness for autistic students with co-occurring conditions, particularly when implemented by frontline staff rather than PT experts. This study addressed these gaps. Five autistic students with co-occurring conditions participated; all had Education, Health and Care Plans and attended a special education school in England, and four completed the study. Instruction focused on identifying equivalent amounts of money using worksheets and flashcards. Two concurrent multiple-baseline across-participant designs with two tiers were used. Instruction employed learn units and frequency-building, delivered by CABAS teachers in sessions lasting on average 20–25 min. Participants improved, with gains maintained post-instruction and small positive effect sizes. Participants performed well on most fluency assessments and demonstrated emerging skill application. Students with substantial differences in academic performance can benefit from PT, and frontline staff without PT expertise can implement it. However, functional relations were not fully demonstrated, feasibility remains a concern, and further research is needed.

1 | Introduction

Mathematics is a crucial academic area and forms a core part of educational curricula worldwide (Bouck and Long 2023; Education Endowment Foundation 2017, 2020). It underpins a range of advanced competencies, such as problem-solving, numeracy, and critical thinking (Mangarin and Caballes 2024). Developing mathematical proficiency is essential for all students, including both typically developing and autistic individuals, some of whom may have co-occurring conditions such as intellectual disability or Attention-Deficit/Hyperactivity Disorder, as it influences postsecondary education, employment opportunities, and everyday life skills, including budgeting, time management, shopping, cooking, and managing personal finances (Bouck and Long 2023; Codding et al. 2024). Despite its importance and benefits, mathematics is a complex domain that can be challenging for all students, whether or not they have a diagnosis, as it requires interactive skills such as fluency in basic mathematical operations, conceptual and procedural knowledge, and text comprehension, among others (Mangarin and Caballes 2024). Moreover, when mathematics is not conditioned as a reinforcer (Maurilus and Greer 2024), the high response effort required to engage with it may lead to avoidance or escape behaviors, and negative emotions, often associated with what is commonly referred to as math anxiety (Mangarin and Caballes 2024). Therefore, motivation and preference for mathematics are important variables that, when not adequately established, can contribute to further difficulties for students. Importantly, difficulties in mathematics are not unique to autistic students or those with other diagnoses (Fuchs et al. 2005; Mangarin and Caballes 2024).

However, certain populations may experience more pronounced or persistent difficulties. Mathematical achievement among autistic individuals has received some attention, with research indicating that they may require additional educational support and often experience greater challenges in mathematics than their typically developing peers (Mayes and Calhoun 2006). Moreover, evidence underscores the need for targeted, individualized interventions to promote mathematical development, particularly for autistic students with co-occurring intellectual disabilities (Park et al. 2020). As a result, in recent years, there has been a growing emphasis on the need for evidence-based procedures designed to enhance students' mathematical proficiency.

However, the literature remains limited, revealing a notable gap in effective, empirically validated mathematical instructional procedures for autistic learners with co-occurring conditions (Fleury et al. 2014). One field that has historically yielded encouraging outcomes in developing mathematical skills is Precision Teaching (PT). This system focuses on measuring the effects of instruction, enabling teachers to make dynamic and strategic decisions about their students' progress (Evans et al. 2021). Although PT is fundamentally a measurement system, it is inherently integrated with instruction, as it is designed to guide data-based instructional decision-making. As a result, when implemented to develop skills, it incorporates various instructional strategies within its framework (Evans et al. 2021).

1.1 | The Precision Teaching Framework

Precision Teachers follow a five-step framework: (a) pinpoint, (b) deliver instruction, (c) chart, (d) decide, and (e) try again (Evans et al. 2021). In the pinpoint stage, they identify target skills using movement cycles (e.g., "writes letter") and learning channel sets (e.g., See-Say; see Table 1), which improve data accuracy (Johnston et al. 2019; Kubina et al. 2024; Kubina and Cooper 2000). During instruction, they flexibly apply teaching strategies, as the focus of PT is not to dictate the strategies used but to offer information on their impact. In charting, they use Standard Celeration Charts (SCCs) to track performance over varying timeframes (e.g., daily, weekly, or monthly) and analyze learning metrics such as celeration, which quantifies the direction and rate of change (e.g., $\times$ 2 or $\div$ 4; Calkin 2005). In decide, they evaluate progress to determine whether to continue, advance, or try again by adjusting instruction.

1.2 | Instructional Approaches Within Precision Teaching

When deciding on which instructional strategies to use, Precision Teachers consider their students' stage of learning. Stages of learning refer to the different steps in the instructional process, where students transition from an acquisition stage to a proficiency stage, and eventually to an application stage (Jimenez et al. 2021). In the acquisition stage, the focus is on developing students' accuracy through untimed activities that allow Precision Teachers to provide confirmatory and corrective feedback depending on their students' needs. Various strategies can be employed, such as self-paced activities, choral responding, and response prompts (e.g., modeling and gestures), to name a few (Cowan et al. 2023; Johnson and Street 2012). Yet although choral responding is particularly effective, it cannot always be used with students with more complex academic needs.

One instructional strategy that could be particularly effective for these students is the learn unit. Learn units, sometimes referred to as discrete trials, represent an interlocking three-term contingency which characterizes a fundamental unit of instruction (Greer and McDonough 1999). They are the primary measure of learning in a CABAS setting and include both pupil and teacher responses. The CABAS (i.e., Comprehensive Application of Behavior Analysis to Schooling) model is a research-driven, system-wide approach to teaching children and young people with and without disabilities (Hawkins and Grant 2026; Singer-Dudek et al. 2010). It applies a pedagogical science to all aspects of teaching and learning, and learn units represent an essential component of this system, each comprising three components: antecedent, behavior, and consequence. Learn units are linked to a restricted-operant paradigm, in which the pace of student responding is influenced by the instructor's presentation of learn units. That is why CABAS teachers strive for a brisk pace of instruction to offer students as many practice opportunities as possible during each instructional period (Du et al. 2026; Ross et al. 2005). Learn units are considered a key component of effective instruction and have been shown to increase correct responses by 4–7 times (Albers and Greer 1991). Therefore, they could be particularly impactful for students at the acquisition stage of learning, where the focus is on developing accurate responses.

Once students improve their accuracy, Precision Teachers transition them to the proficiency stage, where the focus is on developing a natural pace of responding that is accurate and with little hesitation. This is also known as behavioral fluency and has been shown to lead to a series of by-products, namely, endurance, stability, maintenance, and application (Binder 1996). In this stage, frequency building to a performance criterion is employed, allowing students to perform at their natural pace (Kubina and Yurich 2012). Typically, short, timed practice is provided based on students' age, ability, and the task at hand. During this practice, students are presented with one antecedent (e.g., a request to answer all items on a worksheet) and are allowed to continuously respond until the end of the timing. Then students receive confirmatory and corrective feedback. As a result, frequency-building activities are linked to a free-operant paradigm (Bulla et al. 2026). In summary, PT is particularly versatile and open to various instructional procedures that match the students' learning hierarchy.

1.3 | Research Gaps in Precision Teaching

Recently, Bulla et al. (2024) compared restricted and free-operant arrangements and concluded that it is not a matter of which to use, but when each one should be used. Therefore, a combination of acquisition- and proficiency-focused activities appears optimal. Although PT has previously combined these activities, to our knowledge, PT has never employed the learn

TABLE 1 | A description of key terms used in the study.

Movement cycle	Refers to the smallest unit of measurement that captures the beginning, middle, and end of a response. Movement cycles can be repeated, and by measuring them, it is possible to precisely track the occurrence of behavior over time. Examples: Writes a letter, says a word, draws a circle. In this case, each letter written, word said, and circle drawn would be counted as a single response. For more information: (Johnston et al. 2019)
Learning channels set	Refers to how an individual perceives an antecedent and subsequently responds to it. They allow precision teachers to explicitly account for the modality of instruction delivered. Examples: • An activity on the see-say channel set indicates that the individual will see the question and say the answer. • An activity on the hear-write channel set indicates that the individual will hear the question and write the answer. For more information: (Kubina and Cooper 2000; Vostanis et al. 2022)
Standard celeration charts	A family of standardized visual displays that include daily, weekly, monthly and yearly versions. They are a primary tool of precision teachers, and they allow them to evaluate both the absolute and proportional improvements of behavior change across time. In other words, they allow precision teachers to evaluate how much change in behavior has occurred across time, and how substantial that is considering the individual's entry level of ability. Standard celeration charts also enable the calculation of behavioral metrics, such as celeration and bounce, that quantify aspects of behavior change across time. For more information: (Evans et al. 2021; Kubina and Yurich 2012)
Frequency	The count of responses per unit of time, typically reported as count per minute. Examples: • 20 correct responses per 5 min, which would be charted as 4 responses per minute. • 10 requests per hour, which would be charted as 0.17 responses per minute.
Celeration	A behavioral metric used in precision teaching to quantify the trend line created by data on the chart. It allows precision teachers to evaluate the direction of behavior change over time (e.g., increasing, stable, or decreasing), as well as the rate of that change, that is, how quickly behavior increases or decreases. This is an essential metric, as it supports decision-making based not only on performance at a single point in time, but also on overall growth across days of instruction. In other words, celeration measures learning as a process. Celeration is reported using a multiplication symbol (×) to denote acceleration and a division symbol (÷) to denote deceleration. A × 2.00 acceleration is considered the gold standard for many, though not all, instructional programs, as it represents a doubling of performance over time. Celeration is typically calculated weekly, though this depends on the instructional schedule. For more information: (Bulla and Kubina 2026; Kubina and Yurich 2012)
Bounce	A behavioral metric that quantifies variability in performance over time. It is expressed using a multiplication symbol to indicate the degree of variability in the data set. A bounce of × 3 or lower is generally considered an indicator of stability. For more information: (Bulla and Kubina 2026; Kubina and Yurich 2012)
Level change multiplier	A behavioral metric that quantifies the change in average performance across conditions (e.g., baseline vs. instruction). A multiplication symbol indicates an increase in average performance, while a division symbol indicates a decrease. For more information: (Bulla and Kubina 2026; Kubina and Yurich 2012)
Restricted operant	Refers to environmental arrangements in which opportunities to respond are controlled or limited by external factors, such as the instructor's pace during a

(Continues)

	lesson. Example: Learn units require the instructor to present an antecedent, such as a question, before the student can respond. As a result, the student's response rate is influenced by the instructor's pacing. For more information: (Bulla et al. 2024, 2026)
Free operant	Refers to environmental arrangements in which opportunities to respond are not restricted by the instructor, allowing the student to respond at their natural pace. Example: During frequency-building activities, students may be given a multi-page worksheet and asked to complete as many items as possible within a set time (e.g., 1 min). For more information: (Bulla et al. 2024, 2026)
Timing	An instructional strategy in which a student is given an activity and asked to perform continuously until a predetermined duration ends. It is typically used in frequency-building to help students develop their natural response rate. Durations vary depending on the learner's age, current skill level, and the nature of the task.

unit as an acquisition-focused strategy, despite its widespread use in other educational models, such as CABAS, across various skill domains (Baldonado and Greer 2025; Sun and Stropoli 2026; Verdun et al. 2022). Moreover, a recent systematic review by McTiernan et al. (2022) found that 98% of PT participants were typically developing, with only 2% having IDD. Researchers implemented PT in 57% of studies, underscoring the need for evidence on its use by teachers and teaching assistants. Although recent studies (Vostanis et al. 2021; Vostanis et al. 2022) have applied PT with students with IDD, these involved researcher-led sessions with advanced learners. Thus, a gap remains regarding PT's implementation by teachers supporting students with more complex needs.

To that end, this study aimed to address two primary research questions:

a. Can autistic students with co-occurring conditions and substantial differences in academic performance engage in instruction focused on accuracy via the learn unit and fluency via frequency building when delivered within a PT framework?
b. What will be the outcomes of this packaged approach when conducted by teachers, instead of PT researchers, in a special education school?

2 | Method

2.1 | Participants and Setting

Five male pupils, aged 14–17 years old, with a primary diagnosis of autism and co-occurring conditions (see Table 2), all of whom had an Education, Health and Care Plan, participated in the study. These students attended a CABAS special education school in England, rather than a mainstream educational setting.

2.2 | Ethical Approval

Ethical approval was obtained from the University of Kent, and parental consent was secured before the study. Participant assent was monitored continuously, with assent and dissent behaviors recorded each session. Sessions were paused for a 5 min break following dissent. If assent was not indicated, the session ended for the day. Breaks were repeated as needed, and the number of breaks was recorded.

2.3 | Inclusion Criteria

For inclusion in the study, the participants needed to have:

a. A diagnosis of autism.
b. Scored below 85 on the Test of Mathematical Abilities-3.
c. Demonstrated the following skills from the Early Learner Curriculum and Achievement Record:
    - Count from 1 to 20.
    - Add manipulatives and calculate the sum.
    - Add single digits when written.
    - Count with 1:1 correspondence.
    - Name and identify British coins.

2.4 | Materials

2.4.1 | *Assessment Tools*

A series of assessment tools was used primarily for descriptive purposes (see Table 2). The Gilliam Autism Rating Scale–2nd edition (GARS-2; Gilliam 2006) is a 42-item measure of autistic traits, comprising three 14-item subscales: stereotyped behaviors, communication, and social interaction. The Test of Mathematical Abilities–3rd edition (TOMA-3; Brown et al. 2013) is a 145-item test assessing mathematical symbols and concepts, computation, everyday maths, word problems, and attitudes toward maths. The Vineland Adaptive Behavior Scales–2nd edition, Teacher Rating Form (VABS-II TRF) Sparrow et al. 2005) is a 233-item measure of adaptive behavior across communication, daily living skills, socialization, and motor skills, producing a composite score. The Early Learner Curriculum and Achievement Record–2nd edition (ELCAR; Greer et al. 2020) is a CABAS developmental inventory comprising screenings, an achievement record, and a verbal behavior development assessment.

TABLE 2 | Participants' additional diagnostic information and assessment scores at the start of the study.

Participants	Additional diagnostic information	TOMA-3 Standard score	TOMA-3 % Tile	GARS-2 Standard score	VABS-II TRF Standard score	VABS-II TRF % Tile
P1	Oral dyspraxia; social-communication difficulties; OCD.	49	< 1%	98	54	< 1%
P2	Social-communication difficulties.	49	< 1%	100	51	< 1%
P3	ID; speech and language delays.	51	< 1%	100	62	1%
P4	Epilepsy; speech and language delays.	52	< 1%	83	62	1%
P5	Social-communication difficulties.	76	5%	61	92	30%

Note: The VABS-II scores were calculated excluding the motor subdomain. These assessments are meant to provide additional information about the participants' overall abilities and needs. They were not used to evaluate the outcomes of the study.
Abbreviations: ID = intellectual disability, OCD = obsessive-compulsive disorder.

2.4.2 | General Classroom Materials

Participant materials included individualized token boards or wallets containing money, which could be exchanged for items in the class shop. Data sheets and daily graphs were constructed using Microsoft Word. Baseline data sheets included a table for each participant with rows for each session and columns for correct and incorrect responses per timing. Instructional data sheets recorded correct and incorrect responses for both worksheets and flashcards, daily performance goals, assent withdrawals, and breaks taken during each session. Participants were provided with a data sheet and graph for each activity (worksheets and flashcards), which were replaced every 4 weeks.

The data sheet consisted of five columns, one for each day of the week (Monday to Friday), with three rows, one for each timing. The participants' timing graph consisted of the days of the week along the *x*-axis and a ratio *y*-axis to record the number of correct and incorrect responses.

2.4.3 | Materials for Mathematical Practice

Different versions of worksheets, including different exemplars, were created in Microsoft Word and flashcards in Microsoft PowerPoint. The 10 pence (10p) worksheets used a 15.96 × 23.5 cm table with two columns and three rows; the 20 pound (£20) worksheets had two rows, with the right-hand column divided into three. Each row displayed coins or notes that were either equivalent or not equivalent to the target coin/note shown on the left. Flashcards (16.3 × 9.4 cm, laminated) showed coins or notes on a white background that were either equivalent or not equivalent to the target. A visual prompt displayed the question "Are the coins equivalent to 10p?" or "Are the coins equivalent to £20?" in black font (size 48), alongside an image of the target coin/note in a 25.5 × 16.8 cm box. Additional laminated flashcards (15.4 × 9.8 cm) showed "Yes" or "No" in black font (size 72) with a £20 note or 10p coin marked by a green tick or red cross.

2.4.4 | Generalization Materials

Additional materials for generalization assessments included laminated flashcards (17.6 × 9.8 cm) with a purple background and worksheets similar to those used in teaching. In these worksheets, the target exemplar column had an orange background, and different coin and note exemplars were used.

2.5 | Dependent Variable

The mathematical skill assessed was identifying equivalent amounts of money to a specified target value. The first skill was pinpointed as See-Draws line from the target value to the correct amount of coins or notes in a field size of three. The dependent variable was the correct and incorrect matching of the target value to the equivalent value of coins/notes during a 30s timing. The second skill was pinpointed as See-Match Places card on Yes/No pile to indicate whether the amount shown on the flashcard provides the correct sum of money. The dependent variable was the correct and incorrect sorting responses of flashcards depicting differing amounts (i.e., pictures of coins) to either the "Yes" or "No" flashcard according to whether that amount was equivalent to the target value or not during a 30s timing. For both skills, the data were collected after each round of timed practice. Target combinations were divided into two subsets. Each subset targeted three different combinations, and the review subset targeted all six combinations for each skill.

2.6 | Experimental Design

This study employed two concurrent multiple baseline designs across participants with two tiers. Participants were allocated to the different tiers through a simple randomization process by assigning each participant a number and then using a dice roller via www.random.org. Initially, we aimed to employ a three-tier design; however, due to Participant 1 displaying assent withdrawal, he was removed from the study, and two more participants were added in a delayed fashion.

2.7 | Teacher Training

Instruction was delivered by the first three authors, who were trained in the CABAS model for a minimum of 5 years. One teacher was certified as a CABAS Master Teacher, the other was in the process of completing the CABAS Master Teacher rank, and the third teacher was working toward the CABAS Teacher II rank. As part of their CABAS training, all teachers receive extension preparation in behavior analytic instructional practices, including delivering learn units, implementing reinforcement systems, data-based decision-making and collecting data. Specifically, Master Teachers have experience in making rapid, data-based instructional decisions and further learn to implement precise, scripted and replicable teaching procedures. These competencies overlap directly with the skills required to implement the PT procedures. However, they were only partially fluent in frequency-building activities and had no prior experience with the SCC. Therefore, additional meetings were conducted with the last author, a BCBA doctoral-level practitioner and PT researcher. Specifically, three training sessions were held online via MS Teams, during which the last author modeled key components of the frequency-building procedures, explained the SCC, and addressed related questions. Teachers also had the opportunity to review the study materials, discuss their application with the last author, and ask further questions about the procedures.

2.8 | Procedure

2.8.1 | *Baseline*

During baseline, participants were provided with two 30s timings, in a 1:1 format, for each skill and were instructed to perform at their natural pace with no feedback provided. Data were collected over 3 weeks in a staggered manner, as per the experimental design. Participant 1 received 8 baseline sessions, Participant 2 and 3 received 12 sessions, and Participants 4 and 5 received five baseline sessions each.

2.8.2 | *Instruction*

Sessions were delivered in a 1:1 format. At the beginning, participants chose what they wanted to work for. Instruction consisted of goal setting, untimed and timed practice, graphing, and a token economy system (see Table 3).

2.8.2.1 | *Goal Setting.* At the start of each subset's first day, participants completed a 30 s set-criterion timing to establish their baseline performance, with the day's criterion set one response higher. They were expected to meet or exceed this criterion on subsequent practice days. The set-criterion timing occurred only on the first day of a new or review subset. On later days, the criterion increased by one correct response if met or stayed the same if not.

2.8.2.2 | *Untimed Practice.* Untimed practice was conducted for each skill separately but in an alternating sequence. The order of presentation was pre-determined through randomization, using the previously described dice-rolling method. During the untimed practice sessions, participants completed ten See-Draw opportunities and ten See-Match opportunities.

2.8.2.3 | *Timed Practice.* Following each round of untimed practice, participants engaged in a 30 s timing. Untimed and timed practice was repeated until each participant met their daily performance criteria or a maximum of three rounds for each skill. Practice on each subset continued for 10 days before moving to the next.

2.8.2.4 | *Graphing.* At the end of each practice round, participants transferred their data to the datasheet, and at the end of the session, they plotted their best score of the day on the graph provided.

2.8.2.5 | *Token Economy.* Tokens were delivered throughout the session according to a variable ratio schedule of reinforcement (VR 3). At the end of each session, tokens were exchanged for a preferred activity or item from the class shop. The class shop offered a range of options with differing values, with more preferred items assigned higher values. All options were selected based on participants' prior engagement with activities and items identified as highly preferred.

2.9 | Mastery Assessment (ESGA-M)

At the end of the study, brief mastery assessments were conducted. Endurance involved a 90 s timing for each skill using the same materials. Stability involved 30 s timings with distracting but mathematically relevant stimuli (e.g., maths songs, peers practicing maths). Generalization used a novel instructor, setting, and materials. Application involved a 3 min timing with real coins, where participants created target values from provided coins while the instructor reset the pile after each response. Specifically, the instructor presented a variety of different coins to the participant (more than they required) and asked them to make the target value using the coins provided. Once the participant responded, the coins were gathered back by the experimenter and a new selection of coins was presented for the next response. Maintenance involved two 30s timings weekly for 5 weeks, with the best score plotted on the SCC.

2.10 | Interobserver Agreement

Interobserver agreement (IOA) was calculated by the first author, a Board-Certified Behavior Analyst, who independently scored worksheets and session recordings across all conditions. Agreement was calculated using the total count method, comparing teacher and observer counts of correct and incorrect responses for both learn units and timings. The smaller count was divided by the larger for each category, averaged, and multiplied by 100. Session IOA values were then averaged by condition, and overall mean IOA values were calculated per skill targeted (see Table 4).

2.11 | Procedural Fidelity

The last author scored procedural fidelity. The baseline fidelity checklist comprised 10 steps; the instructional checklist, 16; the fluency assessment, 10; and the maintenance fidelity checklist, 7 (see Data Repository section). The checklists were scored by writing "Yes" or "No" for each step. Procedural fidelity was assessed for all participants across all phases of the study (see Table 4).

2.12 | Social Validity

At the end of the study, participants completed a questionnaire (see Data Repository) with up to eight questions about the training. Specifically, the questionnaire assessed participants' perceptions of (a) the clarity of the study procedures, (b) their experiences with key components of the intervention (daily practice, graphing, timing, and video recording), (c) perceived levels of support, and (d) their willingness to participate in similar studies and engage in additional practice. Seven questions used a seven-point Likert scale (1 = strongly disagree, 4 = neutral, 7 = strongly agree), and one was a yes/no question. For P2, only two yes/no questions were included, accompanied by happy and sad faces to support understanding. The instructor read each question aloud, assisted as needed, and asked participants to tick or circle their responses.

TABLE 3 | The study's instructional protocol.

DV 1: See-draws line from the target value to the correct amount of coins or notes in a field size of three.	
Learn unit presentation (untimed practice–acquisition stage) Total learn units delivered: Up to 40	1. Instructor modeled the first few responses (adding coins/notes and drawing a line to the matching amount). 2. Vocal antecedent/SD: "How much altogether?" was presented for the first combination. 3. Participant or instructor wrote the total value next to the combination. 4. Consequence delivered: Confirmatory feedback if correct, corrective feedback if incorrect. 5. Steps 2–4 were repeated for two further rows (three *addition* learn units in total). 6. Vocal antecedent/SD: "Match/draw a line to find the same" or "Find the same" was then presented. 7. Consequence delivered as above for the *drawing-to-match* learn unit (fourth learn unit delivered). 8. Steps 2–7 were repeated for remaining see-draw responses.
Confirmatory feedback	Vocal praise delivered for each correct response; tokens (money or ticks) provided on a variable ratio of 3 schedule. Tokens were exchangeable at the end of the session for a preferred reinforcer (e.g., iPad time).
Corrective feedback	1. For *addition* learn units: Participant vocally prompted to re-add coins/notes and state the correct amount. 2. Antecedent repeated for another response opportunity. 3. For *drawing-to-match* learn units: Instructor used a gestural prompt/pointed to the correct answer. 4. Antecedent re-presented for another attempt. If correct following feedback, no praise was given; if incorrect, correction repeated up to three times before proceeding.
Frequency-building to a performance criterion (timed practice–proficiency stage)	1. Participant completed as many worksheets as possible in 30 s, drawing lines only (no written sums). 2. Timer started; no vocal antecedents given. 3. After 30 s, instructor recorded correct and incorrect responses. 4. Confirmatory or corrective feedback provided. 5. Participant checked if performance criterion was met; if so, data were recorded and graphed. 6. If not, another round of untimed and timed practice followed.
DV2: See-match places card on yes/no pile to indicate whether the amount shown on the flashcard provides the correct sum of money.	
Learn Unit Presentation (Untimed Practice–Acquisition Stage) Total Learn Units Delivered: Up To 20	1. Instructor modeled first few learn units by adding the coins/notes on a flashcard and placing it in the correct yes/no pile. 2. Vocal antecedent/SD: "How much altogether?" or "How much in total?" was presented with one flashcard. 3. Participant vocally stated the total amount. 4. Consequence delivered: Confirmatory if correct, corrective if incorrect (1 *addition* learn unit). 5. Vocal antecedent/SD: "Find the same", "Match the same", or "Sort" was presented. 6. Participant placed the flashcard in the correct yes/no pile according to equivalence with the target value. 7. Consequence delivered as above (1 *matching* learn unit). 8. Steps 2–7 repeated for 10 flashcards.
Confirmatory feedback	Same as for see-draw.
Corrective feedback	1. For *addition* learn units: Participant vocally prompted to re-add correctly; antecedent re-presented.

(Continues)

DV2: See-match places card on yes/no pile to indicate whether the amount shown on the flashcard provides the correct sum of money.	
	2. For *matching* learn units: Instructor used a gestural prompt/ pointed to the correct yes/no pile and verbally corrected the response. 3. Antecedent re-presented for another attempt. If correct following feedback, no praise was given; if incorrect, correction repeated up to three times before proceeding.
Frequency-building to a performance criterion (timed practice–proficiency stage)	1. Participant matched as many flashcards as possible in 30 s. 2. Timer started; no vocal antecedents given. 3. After 30 s, instructor counted correct and incorrect responses within each pile. 4. Confirmatory or corrective feedback provided. 5. Participant checked if performance criterion was met; if so, data were recorded and graphed. 6. If not, another round of untimed and timed practice followed.

Abbreviation: SD = Discriminative Stimulus.

TABLE 4 | Inter-observer agreement and procedural fidelity data.

Phase	IOA sessions assessed (%)	Mean IOA (%) (range)	Fidelity sessions assessed (%)	Mean fidelity (%) (range)
Baseline (worksheets)	P2: 33% P3[a] P4: 60% P5: 40%	100%	P2: 50% P3: 33% P4: 60% P5: 60%	100%
Baseline (flashcards)	P2: 33% P3: 25% P4: 60% P5: 40%	98.43% (93.75%–100%)	P2: 50% P3: 37.5% P4: 60% P5: 60%	100%
Instruction (worksheets)	30%	Untimed: 98.47% (93.87%–100%) Timed: 99.17% (96.67%–100%)	30%	97.65% (95.3%–100%)
Instruction (flashcards)	30[b]%	Untimed: 97.57% (90.27%–100%) Timed: 100%	30%	97.45% (95.3%–100%)
Mastery assessment	100[c]%	Worksheets: 90.13% (77.05%–100%) Flashcards: 93.16% (75%–100%)	100%	100%
Maintenance	40%	Worksheets: 99.37% (97.50%–100%) Flashcards: 100%	40%	100%

Note: Fidelity refers to procedural fidelity. Ranges are presented in parentheses. Untimed and timed refer to practice conditions during instruction.
Abbreviation: IOA = interobserver agreement.
[a]Baseline data for P3 (worksheets) were unavailable due to corrupt scanned files.
[b]For Instruction (Flashcards), P2 subset 2 included 20% of sessions assessed.
[c]Endurance data for P3 during mastery assessment were unavailable due to a video recording fault.

Social validity was also obtained from two school staff members (see Data Repository), who supported data collection from the participants. The questionnaire used the same scale as the participant one, included one open-ended question for additional comments, and assessed the (a) clarity of the study, (b) feasibility of implementation, (c) perceived participant outcomes, (d) acceptability of the intervention duration, and (e) willingness to implement similar instruction in the future, along with opportunities for additional comments.

2.13 | Data Analysis

Data were plotted using PrecisionX, available at https://centralreach.com/products/precisionx/, which generated SCCs and calculated key behavioral metrics: celeration, bounce, and level change multiplier. As shown in Table 5, metrics enable the precise evaluation of student progress (Bulla and Kubina 2026). For example, Table 5 allows for direct comparison of the observed celeration values against the × 2.00 benchmark, which is commonly regarded as the gold standard for progress in Precision Teaching (Johnson et al. 2019, 2021). This information supports more accurate and informed decision-making by teachers regarding student progress. Additionally, baseline-corrected Tau was used, available at https://ktarlow.com/stats/tau/, as a non-parametric effect size (Tarlow 2017). Effect sizes were classified as small (0–0.65), medium (0.66–0.92), or large (0.93–1.00).

3 | Results

This study produced positive effects for all participants, but a functional relation was not established due to the limitations of our experimental design. Nevertheless, all participants demonstrated improvements in their performance that were maintained well when the instruction was completed and produced small yet positive effect sizes (see Figures 1 and 2), except P4 in the flashcards-based activity.

3.1 | Worksheets

During baseline, P2 averaged 3 correct and 10 incorrect responses per min (see Figure 1, first panel). During instruction, P2 gradually improved their performance, averaging 36 correct and 1 incorrect per min in the Review set. During the mastery assessment, they maintained their performance well in the ESG and demonstrated emergent application with 31.25% accuracy (see Figure 3). During maintenance, their performance was maintained above baseline levels with no overlap, averaging 26 correct and 0 incorrect per min. Their average celeration during instruction was encouraging (× 1.99, range: × 1.12–× 3.62), while their bounce indicated stable performance (see Table 5).

During baseline, P3 averaged 19 correct and 3 incorrect per min (see Figure 1, second panel). They gradually improved their performance, averaging 51 correct and 0 incorrect in the review set. Their performance in the ESG was maintained well, and they demonstrated emergent application with 66.67% accuracy. During maintenance, they averaged 36 correct with 0 incorrect per min and no overlap with baseline. Their average celeration was low (× 1.33, range: × 1.08–× 1.51), while their bounce suggested stable performance.

TABLE 5 | Participants' learning metrics.

		Worksheets				Flashcards			
		Celeration		Bounce		Celeration		Bounce	
Participants	Phase	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect
P2	Baseline	÷ 1.03	÷ 1.02	× 2.30	× 3.30	× 2.33	× 1.27	× 6.10	× 6.20
	Subset 1	× 3.62	÷ 1.18	× 2.30	× 3.30	× 2.10	× 2.76	× 1.80	× 2.40
	Subset 2	× 1.12	÷ 3.23	× 1.80	× 3.40	× 1.18	÷ 1.18	× 1.60	× 3.00
	Review set	× 1.23	÷ 2.06	× 1.70	× 5.10	× 1.08	÷ 3.19	× 1.50	× 3.00
	Maintenance	÷ 1.09	× 1.00	× 1.40	× 1.00	× 1.06	× 1.70	× 1.20	× 2.10
P3	Baseline	× 1.05	× 1.10	× 2.30	× 4.80	× 1.12	÷ 1.33	× 1.70	× 4.90
	Subset 1	× 1.39	× 1.00	× 1.40	× 1.00	× 1.06	÷ 1.19	× 1.10	× 1.70
	Subset 2	× 1.51	÷ 1.37	× 1.50	× 2.00	× 1.00	× 1.00	× 1.30	× 1.00
	Review set	× 1.08	× 1.00	× 1.50	× 1.00	× 1.11	× 1.00	× 1.30	× 1.00
	Maintenance	÷ 1.02	× 1.00	× 1.30	× 1.00	÷ 1.04	× 1.00	× 1.30	× 1.00
P4	Baseline	× 1.17	÷ 1.18	× 2.50	× 2.20	× 1.10	× 1.09	× 1.10	× 3.60
	Subset 1	× 1.40	× 1.00	× 1.40	× 1.00	× 1.10	× 1.00	× 1.30	× 1.00
	Subset 2	× 1.03	÷ 1.03	× 1.50	× 1.70	× 1.01	× 1.00	× 1.40	× 1.00
	Review set	× 1.07	× 1.04	× 1.20	× 2.10	× 1.08	× 1.08	× 1.50	× 1.80
	Maintenance	× 1.04	× 1.00	× 1.40	× 1.00	÷ 1.06	× 1.00	× 1.50	× 1.00
P5	Baseline	× 1.26	× 1.12	× 1.50	× 2.00	÷ 1.12	× 1.00	× 1.70	× 1.00
	Subset 1	× 1.36	× 1.13	× 1.60	× 1.80	× 1.27	÷ 1.09	× 1.30	× 1.80
	Subset 2	× 1.50	× 1.00	× 1.60	× 1.00	× 1.35	÷ 1.05	× 1.30	× 1.80
	Review set	× 1.06	× 1.00	× 1.60	× 1.00	× 1.09	× 1.13	× 1.30	× 1.80
	Maintenance	÷ 1.09	× 1.00	× 1.10	× 1.00	÷ 1.11	× 1.00	× 1.20	× 1.00

Note: A multiplication symbol (×) indicates an acceleration in performance over time, while a division symbol (÷) denotes a deceleration. PrecisionX was used to calculate these metrics. Baseline and instructional metrics (i.e., Subsets and Review sets) were derived from the daily per-minute Standard Celeration Chart, whereas maintenance metrics were calculated using the weekly per-minute Standard Celeration Chart. Celeration values close to × 2 or ÷ 2 indicate substantial changes in growth, representing a doubling (100% acceleration) or halving (50% deceleration) of performance, respectively. A × 2 acceleration of correct responses during instruction is considered the gold standard in Precision Teaching for most, though not all, skills (see Johnson et al. 2021 for a detailed discussion on minimum celeration expectations). Therefore, celeration values for correct responses close to or exceeding × 2 may indicate particularly effective instruction. Bounce values are always presented using a multiplication symbol. Values up to × 3 suggest stable performance. A celeration value of × 1.00 indicates no acceleration or deceleration, while a bounce value of × 1.00 reflects no variability in performance.

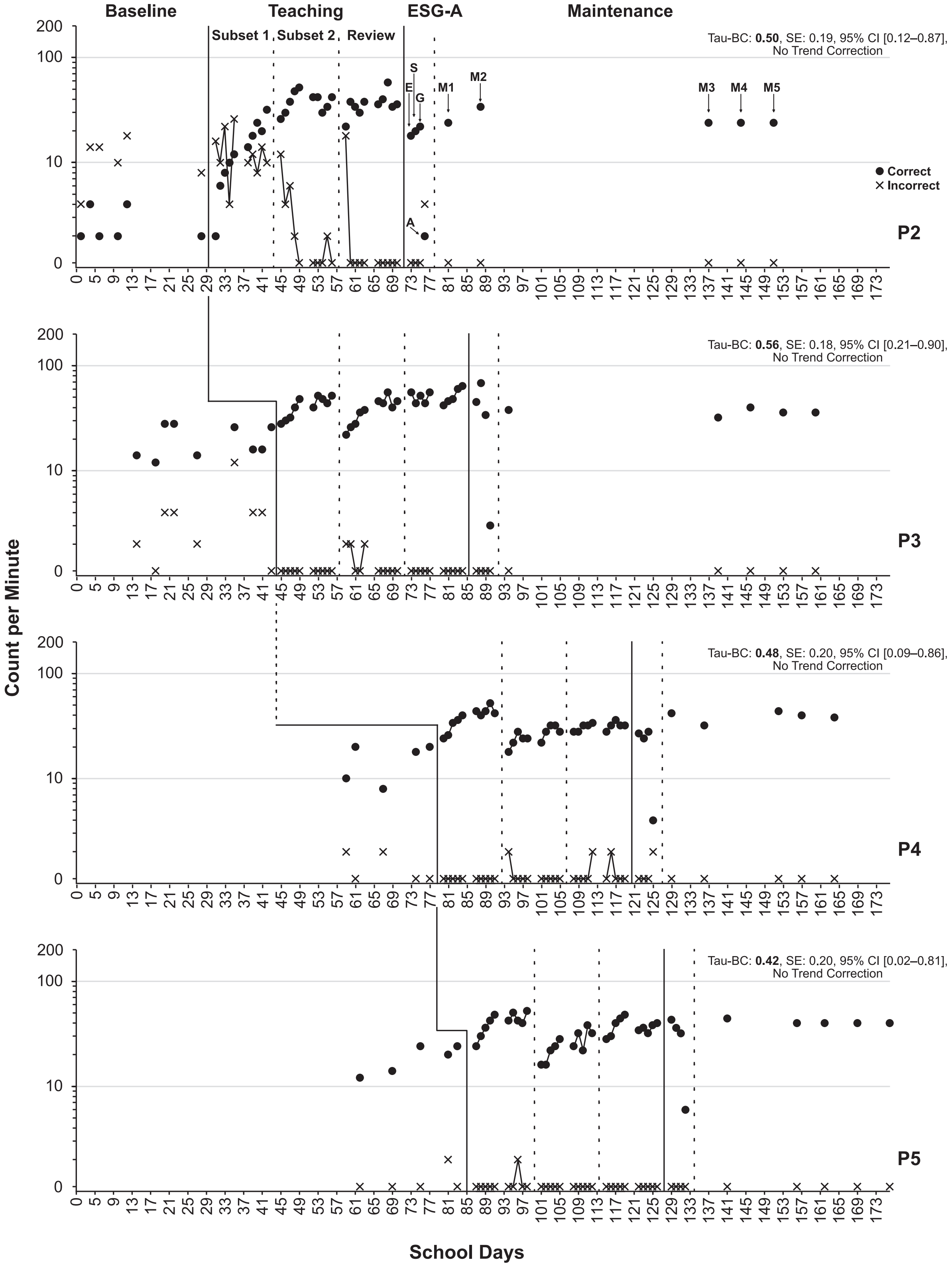

FIGURE 1 | Participants' performance on worksheet-based practice. P2 and P3 were introduced to the baseline phase concurrently, while P4 and P5 were also introduced concurrently but at a later stage. Baseline Corrected Tau was used as the effect size. The ESG-A assessment was conducted on the same day but is presented separately to aid visual analysis. Each datum point represents an individual assessment. Zeros were plotted following the conventions of Standard Celeration Charting. We strongly encourage readers to access the Standard Celeration Charts in the open-data repository to facilitate a more thorough data evaluation. A = application, CI = confidence interval, E = endurance, G = generalization, M1–M5 = maintenance sessions 1–5, S = stability, SE = standard error.

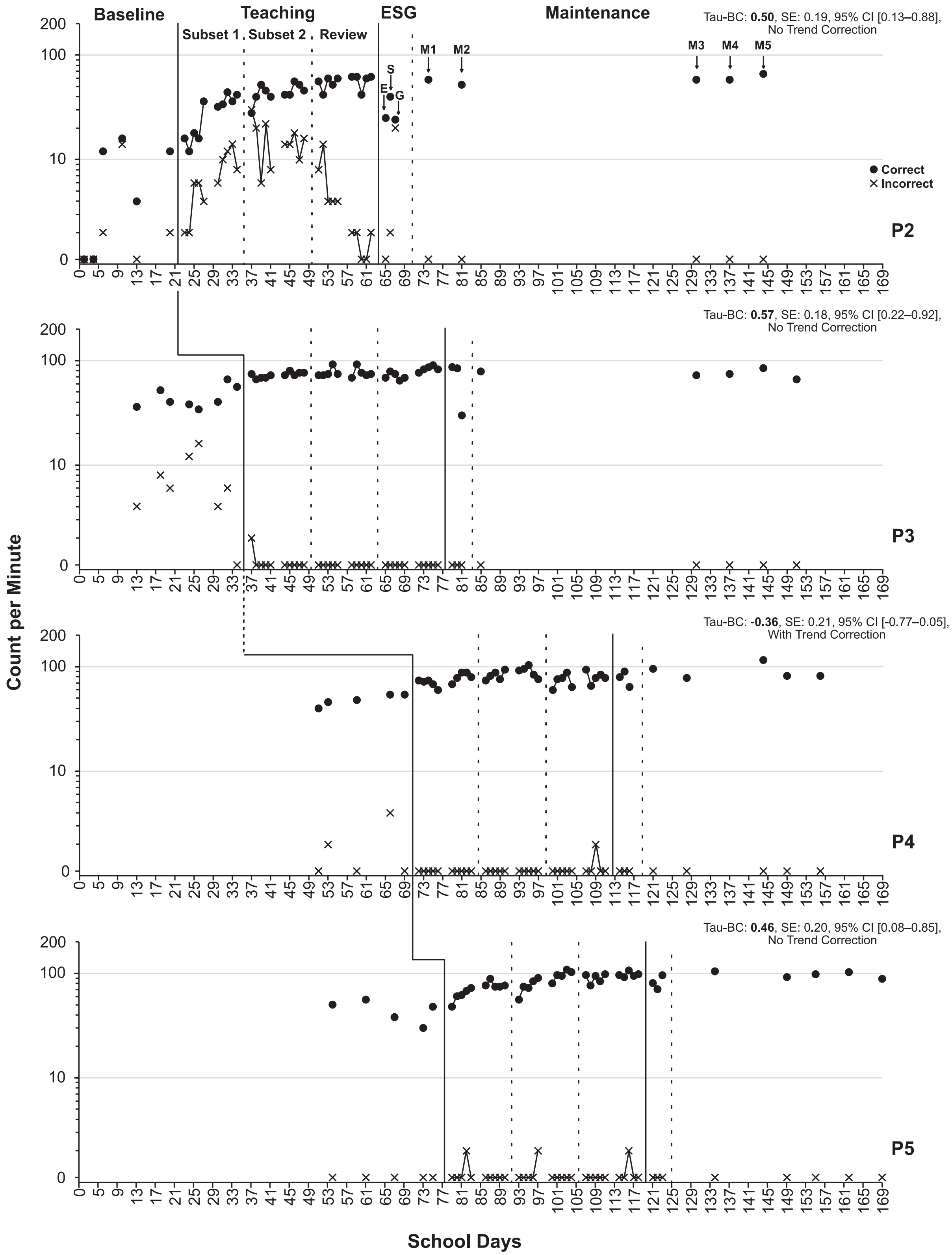

FIGURE 2 | Participants' performance on flashcards-based practice. P2 and P3 were introduced to the baseline phase concurrently, while P4 and P5 were also introduced concurrently but at a later stage. Baseline Corrected Tau was used as the effect size. The ESG assessment was conducted on the same day but is presented separately to aid visual analysis. Each datum point represents an individual assessment. Zeros were plotted following the conventions of Standard Celeration Charting. Application was assessed once, and the data are presented in Figure 1. We strongly encourage readers to access the Standard Celeration Charts in the open-data repository to facilitate a more thorough data evaluation. We chose the term "flashcard-based practice" because participants did not engage in a Say All Fast a Minute Each Day (SAFMEDS) task, a common activity in Precision Teaching, to ensure clarity. CI = confidence interval, E = endurance, G = generalization, M1–M5 = maintenance sessions 1–5, S = stability, SE = standard error.

P4 had an average of 14 correct and 1 incorrect in baseline (see Figure 1, third panel). They gradually improved, reaching 31 correct and 1 incorrect in the review set. Their performance in the ESG was maintained well, and they demonstrated emergent application with 68.75% accuracy. During maintenance, they averaged 39 correct and 0 incorrect, with no overlap with baseline. Their average celeration was low (× 1.17, range: × 1.03–× 1.40), while their bounce suggested stable performance.

P5 averaged 18 correct and 1 incorrect per min in baseline (see Figure 1, fourth panel). They gradually improved their performance, reaching 37 correct with 0 incorrect in the review set. Their performance in the ESG was maintained well, and they demonstrated moderate application with 100% accuracy. During maintenance, they averaged 41 correct and 0 incorrect, with no overlap with baseline. Their average celeration was low (× 1.31, range: × 1.06–× 1.50), while their bounce indicated stability.

3.2 | Flashcards

During baseline, P2 averaged 5 correct and 2 incorrect responses per min (see Figure 2, first panel). During instruction, P2 gradually improved their performance, averaging 55 correct and 3 incorrect per min in the Review set. During the mastery assessment, they maintained their performance well in the endurance and stability assessment, but there was an increase in incorrect responses in the generalization assessment. During maintenance, their performance was maintained above baseline levels with no overlap, averaging 58 correct and 2 incorrect per min. Their average celeration during instruction was low (× 1.45, range: × 1.08–× 2.10), while their bounce indicated stable performance (see Table 5).

During baseline, P3 averaged 44 correct and 6 incorrect per min (see Figure 2, second panel). They gradually improved their performance, averaging 76 correct and 0 incorrect in the review set. Their performance in the mastery assessment was maintained well in the endurance and stability assessments but dropped during the generalization assessment. During maintenance, they averaged 75 correct with 0 incorrect per min and 20% overlap with baseline. Their average celeration was low (× 1.06, range: × 1.00–× 1.11), while their bounce suggested stable performance.

P4 had an average of 48 correct and 2 incorrect in baseline (see Figure 2, third panel). They gradually improved, reaching 76 correct and 1 incorrect in the review set. Their performance in the ESG was maintained well. During maintenance, they averaged 90 correct and 0 incorrect, with no overlap with baseline. Their average celeration was low (× 1.06, range: × 1.01–× 1.10), while their bounce suggested stable performance.

P5 averaged 43 correct and 0 incorrect per min in baseline (see Figure 2, fourth panel). They gradually improved their performance, reaching 93 correct with 1 incorrect in the review set. Their performance in the ESG was maintained well, and they demonstrated moderate application. During maintenance, they averaged 97 correct and 0 incorrect, with no overlap with baseline. Their average celeration was low (× 1.24, range: × 1.09–× 1.35), while their bounce indicated stability.

3.3 | Overall Improvement

P2's performance in worksheets from baseline to maintenance showed an average increase in correct responding by a factor of × 10.21 and a decrease in incorrect responding by ÷ 10.20 (see Figure 4). In flashcards, their average correct responses increased by × 12.74, and incorrect responses decreased by ÷ 1.12 (see Figure 5). P3's corrects in worksheets increased by × 1.91, and incorrects decreased by ÷ 2.84. In flashcards,

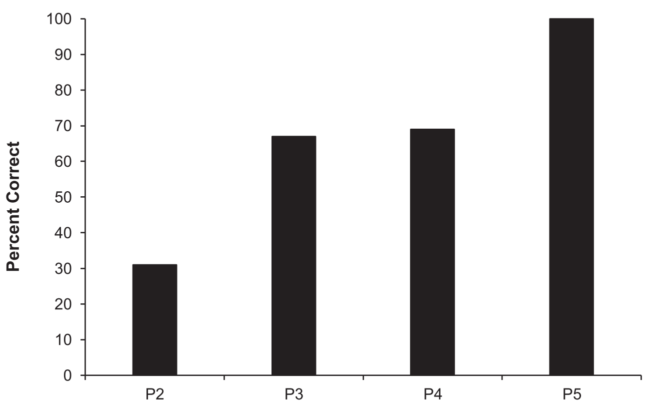

FIGURE 3 | Participants' accuracy in the application assessment. Participants received no training in calculating sums using actual coins. This task was included during the mastery assessment to evaluate whether participants could recombine their taught skills (via worksheets and flashcards) to perform a more complex task. The assessment was delivered once, with a 3-min timing. Participants' frequency of correct and incorrect responses during this task is presented in Figure 1. The ratio of correct to incorrect responses was used to calculate participants' accuracy percentage when performing the task.

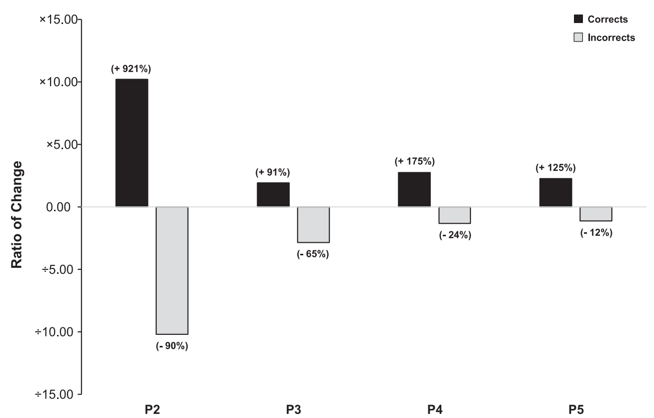

FIGURE 4 | Participants' overall improvement in correct and incorrect responses during worksheet-based practice, from baseline to maintenance conditions. The level change multiplier was calculated by comparing the average number of correct and incorrect responses between baseline and maintenance conditions. The multiplication symbol indicates an increase, while the division symbol indicates a decrease. The resulting ratios were converted into percentages to improve clarity. Percentages with the addition symbol indicate an increase, and percentages with the subtraction symbol indicate a decrease. Please note that some percentages may exceed 100%, which is technically incorrect but is used here to illustrate the improvement more clearly.

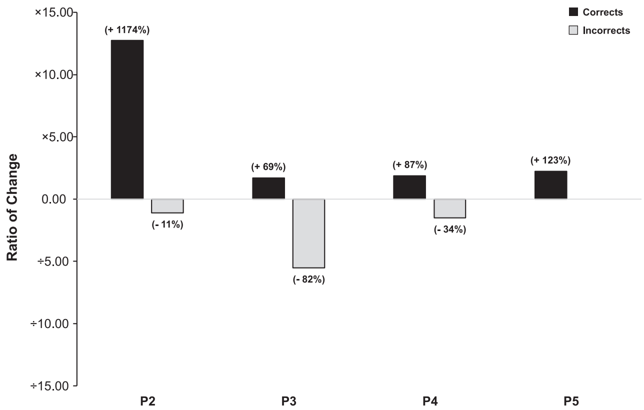

FIGURE 5 | Participants' overall improvement in correct and incorrect responses during flaschards-based practice, from baseline to maintenance conditions. The level change multiplier was calculated by comparing the average number of correct and incorrect responses between baseline and maintenance conditions. The multiplication symbol indicates an increase, while the division symbol indicates a decrease. The resulting ratios were converted into percentages to improve clarity. Percentages with the addition symbol indicate an increase, and percentages with the subtraction symbol indicate a decrease. Please note that some percentages may exceed 100%, which is technically incorrect but is used here to illustrate the improvement more clearly.

corrects increased by × 1.69 and incorrects decreased by ÷ 5.53. P4's corrects in worksheets increased by × 2.75, and incorrects decreased by ÷ 1.32. In flashcards, corrects increased by × 1.87 and incorrects decreased by ÷ 1.51. P5's corrects increased by × 2.25, and incorrects decreased by ÷ 1.14. In flashcards, corrects increased by × 2.23, while incorrects remained at zero.

3.4 | Intensity

P2's sessions lasted, on average, 25 min (range: 8–67), and they completed 36 timings with worksheets and 50 with flashcards. P3's sessions lasted, on average, 20 min (range: 7–40), and they completed 65 timings with worksheets and 72 with flashcards. P4's sessions lasted, on average, 20 min (range: 10–34), and they completed 66 and 68 timings, respectively. P5's sessions lasted 22 min (range: 8–43), and they completed 68 and 71 timings, respectively.

3.5 | Social Validity

P3, P4, and P5 reported a clear understanding of the study and high enjoyment of daily money worksheets and graphing their data, with P2 also enjoying participation. P4 and P5 liked the time limit, while P3 was neutral. P3 and P5 were happy to participate in future studies, while P4 was not. All three felt well supported, and all except P4 were willing to complete worksheets as practice.

Two CABAS teachers provided social validity feedback, indicating mixed outcomes. In terms of positive aspects, both reported improvements in participants' daily practice and response rates, suggesting perceived benefits in instructional outcomes. In addition, procedures were described as reinforcing. In contrast, several aspects were viewed less favorably. Both teachers reported only a neutral understanding of the study and described the training sessions as not straightforward. Ratings of procedural acceptability were variable, with one teacher rating session duration as "not at all acceptable" and the other as neutral. Perceived instructional effectiveness also differed, and only one teacher indicated willingness to participate in similar PT-based instruction in the future. Additional comments further highlighted concerns regarding time demands, particularly when untimed and timed practice had to be repeated multiple times. Overall, these findings suggest that while instructional outcomes were viewed positively, procedural features, especially efficiency and complexity, may have limited overall acceptability.

4 | Discussion

This study aimed to answer two research questions. First, can autistic students with co-occurring conditions engage in instruction focused on accuracy via the learn unit and fluency via frequency building when delivered within a PT framework? Second, what will be the outcomes of this packaged approach when implemented by teachers, instead of PT researchers, in a special education school?

4.1 | First Research Question

Regarding the first research question, the results are encouraging but tentative due to the lack of a clear functional relation. Participants displayed substantial differences in academic performance, and only P5 was a more independent learner. Despite this, all participants could engage successfully with the learn units and the frequency-building activities and graph their data in simplified graphs in line with PT conventions. Not only that, but participants also managed to perform at considerably higher frequencies than baseline, demonstrating the potential to improve their skills when they receive instruction tailored to different stages of learning (Bulla et al. 2024; Jimenez et al. 2021). Although instruction was intensive, only one participant (P1) withdrew their assent and was subsequently removed from the study, demonstrating that students with additional needs can participate and enjoy even more engaging forms of instruction, as demonstrated by the duration of sessions and the number of timings they completed.

Moreover, participants performed well in the mastery assessments, demonstrating the by-products of fluency in most cases. This finding is significant, as PT researchers have long advocated for mastery evaluations extending beyond the acquisition stage, which focuses on accuracy (percent correct), to include fluency outcomes measured by response frequency (Evans et al. 2021). Our findings add to the existing PT literature, demonstrating that participants can engage for longer durations (endurance), in the presence of distractions (stability), using different instructors and materials (generalization; Fabrizio and Moors 2003). Therefore, our findings, although tentative, suggest that a more thorough evaluation of mastery might be necessary to help teachers determine whether their students are ready to progress to the next curricular step. In other words, because participants demonstrated endurance, stability, generalization, and emergent application during the mastery assessments, these findings tentatively suggest that incorporating response frequency alongside accuracy and explicitly evaluating fluency by-products may better support teachers in determining whether students are ready to progress within the curriculum. This is particularly important for autistic students with co-occurring conditions and additional academic needs, as in many cases, they progress through the curriculum without having established fluency in the necessary prerequisite skills. As a result, these students struggle to keep up with the content due to what Precision Teachers refer to as cumulative dysfluency (McDowell and Keenan 2001).

Regarding our participants' ability to demonstrate application, the results were encouraging but fell short of optimal. Although participants were not explicitly instructed to handle or use real money, some demonstrated the emergence of this ability without direct training. However, only Participant 5 reached 100% accuracy, and none demonstrated fluency, suggesting missing prerequisite skills. Importantly, although participants demonstrated prerequisite skills, such as counting 1–20 and counting with 1:1 correspondence, to be included in the study, it is possible that additional skills were not fluent in their repertoire. These may have included sorting coins and notes, engaging in skip counting (e.g., counting by 2, 3, or 5s), performing addition involving a two-digit number (such as 10p + 5p), and estimating quantities, among others (Johnson et al. 2021; Stein et al. 2018). Precision Teachers often use the element-compound analysis to identify and train prerequisite (element) skills before targeting complex (compound) skills (Johnson et al. 2021; Kapoor et al. 2023). This study did not conduct such an analysis or collect data to determine which element skills were lacking; therefore, the findings should be interpreted cautiously.

Overall, despite its small scale and limitations, this study addresses a gap in the PT literature highlighted by McTiernan et al. (2022). Most studies have focused on typically developing students or those with a diagnosis who can access academic skills with relative ease (Vostanis et al. 2021; Vostanis et al. 2022). This study included students with co-occurring conditions and substantial differences in academic performance, demonstrating that they could participate successfully in lessons. What is more, it is the first study to combine the learn unit that is widely used in CABAS schools with frequency-building activities within a PT framework. This combination has two implications. First, PT has a rich history of being combined with Direct Instruction (Desjardins and Slocum 1993; Morrell et al. 1995). As a result, untimed activities typically include choral responding and the model-lead-test method followed by frequency-building (Johnson and Layng 1992). In this study, a different combination was used, namely the learn unit and frequency-building. This combination could prove more accessible to students in special education who may struggle to engage in choral responding and require more individualized instruction, as provided in the learn unit's discrete trial format (Greer and McDonough 1999).

Second, Burns et al. (2010) emphasized the importance of tailoring instruction to students' needs, a point reinforced by McTiernan et al. (2021), who found that the success of frequency-building and PT in maths instruction depended on whether students required more acquisition- or fluency-focused practice. We propose that integrating learn units with frequency-building may efficiently target both stages within instruction, though further data are needed to confirm this.

4.2 | Second Research Question

This study also yielded interesting findings regarding the second research question. McTiernan et al. (2022), highlighted the need for more research by front-line staff. This study took a first step toward addressing that research gap. Instruction was delivered by CABAS teachers in a special education school with no prior training in PT. The last author delivered PT training before the study and provided ongoing feedback during procedural fidelity checks. Although fidelity was high (95.3%–100%), the checklists covered only major steps. Minor errors included reversing charting conventions (using crosses for corrects) and occasional deviations from step 8, where teachers ended practice early when students met the correct-response criterion but still exceeded two incorrects.

This indicates that certain aspects of PT may require additional training to ensure procedural fidelity. Given that instructional intensity can moderate frequency-building outcomes (McTiernan et al. 2021; Mellard et al. 2010), Precision Teachers supporting others should take care to implement these elements accurately to optimize instructional outcomes. Most importantly, the study gathered essential evidence regarding PT's acceptability. While the students generally seemed to enjoy the procedures, the teachers had opposing views. The fact that one teacher rated the instruction's duration as unacceptable suggests that more work is needed to find ways to make PT more feasible in educational settings that do not widely use PT in their provision.

4.3 | Limitations

This study had various noteworthy limitations. First, we initially aimed to implement a concurrent multiple baseline design with three tiers, but one participant withdrew, requiring the recruitment of two additional participants. This resulted in two concurrent multiple baseline designs with two tiers, limiting our ability to identify a clear functional relation. Second, some baselines showed minimally ascending trends. While more stable trends would be ideal, instruction was introduced to accommodate school holidays and student availability. Effect size calculations indicated only one baseline required correction, suggesting the slight increase likely had minimal impact. The study included four participants, a reasonable sample for single-case designs, but further replications are needed to assess external validity. Third, although a token economy was employed, we did not conduct formal preference assessments. Therefore, participants' motivation might have also impacted the study's outcomes. Finally, to protect internal validity, we capped timings at three, whereas in PT, practice would depend on progress. This approach is what sets PT apart, as it allows teachers to receive immediate, ongoing feedback within and across sessions (Johnson et al. 2021).

4.4 | Future Directions

Reflecting on this study's limitations, we propose several directions for future research. First, to minimize practice effects associated with repeated exposure to baseline conditions during school holidays and scheduling conflicts, screening probes should be conducted at the outset of the study. Second, to address attrition, researchers should aim to recruit at least five participants or consider alternative experimental designs instead of a multiple baseline across participants. Third, for individuals with greater academic needs, a more thorough evaluation of prerequisite and related skills is recommended, with these skills targeted prior to introducing primary practice conditions. Fourth, although participants provided assent to the procedures (except for P1, who was withdrawn), we recommend using more formal preference assessments to ensure motivation is maintained across instructional sessions. Finally, based on teachers' social validity reports, researchers should explore ways to improve efficiency, such as implementing a three-day-per-week schedule rather than daily sessions.

4.5 | Conclusion

This study sought to address two key research gaps recently identified by McTiernan et al. (2022). Despite its limitations, the study represents an initial step toward bridging both gaps. It also introduces an innovative integration of learn unit instruction and frequency-building activities within a PT framework, specifically tailored for autistic students with co-occurring conditions.

Funding

The authors have nothing to report.

Conflicts of Interest

The first three authors were employed by the school in which the study was conducted. To mitigate potential conflicts of interest and support objective analysis, responsibility for data analysis was assigned to the final author, who was independent of the school. The authors declare no other conflicts of interest.

Data Availability Statement

The data that support the findings of this study (including the SCCs, social validity questionnaires, datasheets, IOA calculations, and fidelity checklists) are openly available in Figshare at: https://doi.org/10.6084/m9.figshare.c.8512248.

References

Albers, A. E., and R. D. Greer. 1991. "Is the Three-Term Contingency Trial a Predictor of Effective Instruction?" *Journal of Behavioral Education* 1, no. 3: 337–354. https://doi.org/10.1007/BF00947188.

Baldonado, L., and R. D. Greer. 2025. "Increased Complexity in Incidental Bidirectional Naming Stimulus Control Enhances Reading in First Graders." *Psychological Record* 75, no. 3: 389–414. https://doi.org/10.1007/s40732-025-00644-1.

Binder, C. 1996. "Behavioral Fluency: Evolution of a New Paradigm." *Behavior Analyst* 19, no. 2: 163–197. https://doi.org/10.1007/BF03393163.

Bouck, E. C., and H. Long. 2023. "Academic Mathematics Instruction and Intervention for Students With Mild Intellectual Disability: An Updated Review." *Education and Training in Autism and Developmental Disabilities* 58, no. 2: 144–161. https://doi.org/10.1177/215416472305800203.

Brown, V. L., M. E. Cronin, and D. P. Bryant. 2013. Test of Mathematical Abilities. 3rd ed. Pro-ED.

Bulla, A., J. Wertalik, and L. Yakabovits. 2024. "Comparing the Effectiveness of Restricted-Operant and Free-Operant Teaching Arrangements on Measures of Acquisition and Fluency Outcomes." *Single Case in the Social Sciences* 1, no. 1: 15–35. https://doi.org/10.15763/issn.1936-9298.2024.1.1.15-35.

Bulla, A. J., A. S. Bradley, J. L. Wertalik, and A. Sullivan. 2026. "A Systematic Replication Investigating the Efficiency and Effectiveness of Restricted- and Free-Operant Programming." *Journal of Applied Behavior Analysis* 59, no. 1: e70048. https://doi.org/10.1002/jaba.70048.

Bulla, A. J., and R. M. Kubina. 2026. "The Use and Benefits of the Standard Celeration Chart in Single Case Research." *Journal of Behavioral Education*. https://doi.org/10.1007/s10864-026-09617-y.

Burns, M. K., D. A. Klingbeil, and J. Ysseldyke. 2010. "The Effects of Technology-Enhanced Formative Evaluation on Student Performance on State Accountability Math Tests: TEFE for Math." *Psychology in the Schools* 47, no. 6: 582–591. https://doi.org/10.1002/pits.20492.

Calkin, A. B. 2005. "Precision Teaching: The Standard Celeration Charts." *Behavior Analyst Today* 6, no. 4: 207–215. https://doi.org/10.1037/h0100073.

Codding, R. S., A. VanDerHeyden, and R. Chehayeb. 2024. "Using Data to Intensify Math Instruction: An Evaluation of the Instructional Hierarchy." *Remedial and Special Education* 45, no. 3: 173–188. https://doi.org/10.1177/07419325231194354.

Cowan, L. S., D. C. Lerman, K. L. Berdeaux, A. H. Prell, and N. Chen. 2023. "A Decision-Making Tool for Evaluating and Selecting Prompting Strategies." *Behavior Analysis in Practice* 16, no. 2: 459–474. https://doi.org/10.1007/s40617-022-00722-8.

Desjardins, A., and T. A. Slocum. 1993. "Integrating Precision Teaching and Direct Instruction." *Journal of Precision Teaching* 10, no. 2: 20–24.

Du, L., R. Nuzzolo, and S. Buttigieg. 2026. "Evaluating the Efficiency of Behavioural Skills Training and the Teacher Performance Rate and Accuracy Measure." *Tizard Learning Disability Review*: 1–22. https://doi.org/10.1108/TLDR-11-2025-0071.

Education Endowment Foundation. 2017. Improving Mathematics in Key Stages 2 and 3. https://d2tic4wvo1iusb.cloudfront.net/production/eef-guidance-reports/maths-ks-2-3/EEF-Improving-Mathematics-in-Key-Stages-2-and-3-2022-Update.pdf.

Education Endowment Foundation. 2020. Improving Mathematics in the Early Years and Key Stage 1. https://d2tic4wvo1iusb.cloudfront.net/production/eef-guidance-reports/early-maths/EEF_Maths_EY_KS1_Guidance_Report.pdf.

Evans, A. L., A. J. Bulla, and A. R. Kieta. 2021. "The Precision Teaching System: A Synthesized Definition, Concept Analysis, and Process." *Behavior Analysis in Practice* 14, no. 3: 559–576. https://doi.org/10.1007/s40617-020-00502-2.

Fabrizio, M. A., and A. L. Moors. 2003. "Evaluating Mastery: Measuring Instructional Outcomes for Children With Autism." *European Journal of Behavior Analysis* 4, no. 1–2: 23–36. https://doi.org/10.1080/15021149.2003.11434213.

Fleury, V. P., S. Hedges, K. Hume, et al. 2014. "Addressing the Academic Needs of Adolescents With Autism Spectrum Disorder in Secondary Education." *Remedial and Special Education* 35, no. 2: 68–79. https://doi.org/10.1177/0741932513518823.

Fuchs, L. S., D. L. Compton, D. Fuchs, K. Paulsen, J. D. Bryant, and C. L. Hamlett. 2005. "The Prevention, Identification, and Cognitive Determinants of Math Difficulty." *Journal of Educational Psychology* 97, no. 3: 493–513. https://doi.org/10.1037/0022-0663.97.3.493.

Gilliam, J. E. 2006. Gilliam Autism Rating Scale. 2nd ed. Pro-Ed.

Greer, R. D., and S. H. McDonough. 1999. "Is the Learn Unit a Fundamental Measure of Pedagogy?" *Behavior Analyst* 22, no. 1: 5–16. https://doi.org/10.1007/BF03391973.

Greer, R. D., J. Speckman, J. Singer-Dudek, et al. 2020. "Early Learner Curriculum and Achievement Record (ELCAR): A CABAS Developmental Inventory." CABAS and the Fred S. Keller School. 2nd ed.

Hawkins, E., and K. Grant. 2026. "Reflecting on 25 Years of the Comprehensive Application of Behaviour Analysis to Schooling (CABAS) Model in the UK." *Tizard Learning Disability Review*: 1–11. https://doi.org/10.1108/TLDR-11-2025-0068.

Jimenez, B., J. Root, J. Shurr, and E. C. Bouck. 2021. "Using the Four Stages of Learning to Assess, Set Goals, and Instruct." *TEACHING Exceptional Children* 56, no. 6: 452–461. https://doi.org/10.1177/00400599211054873.

Johnson, K., and E. M. Street. 2012. "From the Laboratory to the Field and Back Again: Morningside Academy's 32 Years of Improving Students' Academic Performance." *Behavior Analyst Today* 13, no. 1: 20–40. https://doi.org/10.1037/h0100715.

Johnson, K., E. M. Street, A. R. Kieta, and J. K. Robbins. 2021. *The Morningside Model of Generative Instruction: Building a Bridge Between Skills and Inquiry Teaching*. Sloan Publishing.

Johnson, K., E. M. Street, and M. D. Neely. 2019. "Precision Teaching: Fluency and Agility." In *Precision Teaching—A Practical Science of Education*, edited by N. G. Haring, M. S. White, and M. D. Neely, 127–142. Sloan Publishing.

Johnson, K. R., and T. J. Layng. 1992. "Breaking the Structuralist Barrier: Literacy and Numeracy With Fluency." *American Psychologist* 47, no. 11: 1475–1490. https://doi.org/10.1037/0003-066X.47.11.1475.

Johnston, J. M., H. S. Pennypacker, and G. Green. 2019. *Strategies and Tactics of Behavioral Research and Practice*. 4th ed. Routledge.

Kapoor, G., A. Vostanis, S. Mejía-Buenaño, and P. E. Langdon. 2023. "Using Precision Teaching to Improve Typically Developing Student's Mathematical Skills via Teleconferencing." *Journal of Behavioral Education* 34, no. 1: 109–138. https://doi.org/10.1007/s10864-023-09520-w.

Kubina, R. M., and J. O. Cooper. 2000. "Changing Learning Channels: An Efficient Strategy to Facilitate Instruction and Learning." *Intervention in School and Clinic* 35, no. 3: 161–166. https://doi.org/10.1177/105345120003500306.

Kubina, R. M., M. Halkowski, K. K. L. Yurich, K. Ghorm, and N. M. Healy. 2024. "Comparing the Detection Accuracy of Operational Definitions and Pinpoints." *Journal of Behavioral Education* 33, no. 1: 142–162. https://doi.org/10.1007/s10864-022-09485-2.

Kubina, R. M., Jr., and K. K. L. Yurich. 2012. *The Precision Teaching Book*. Greatness Achieved.

Mangarin, R. A., and D. O. Caballes. 2024. "Difficulties in Learning Mathematics: A Systematic Review." *International Journal of Research and Scientific Innovation* XI, no. IX: 401–405. https://doi.org/10.51244/IJRSI.2024.1109037.

Maurilus, E. N., and R. D. Greer. 2024. "The Effect of the Establishment of Preference for Math on Rate of Learning for Pre-Kindergarten Students." *Behavior and Social Issues* 33, no. 1: 339–363. https://doi.org/10.1007/s42822-024-00164-7.

Mayes, S. D., and S. L. Calhoun. 2006. "Frequency of Reading, Math, and Writing Disabilities in Children With Clinical Disorders." *Learning and Individual Differences* 16, no. 2: 145–157. https://doi.org/10.1016/j.lindif.2005.07.004.

McDowell, C., and M. Keenan. 2001. "Cumulative Dysfluency: Still Evident in Our Classrooms, Despite What We Know." *Journal of Precision Teaching and Celeration* 17, no. 2: 1–6.

McTiernan, A., J. Holloway, O. Healy, and R. M. Kubina. 2021. "An Investigation of Moderators of a Precision Teaching and Frequency Building Intervention." *European Journal of Behavior Analysis* 22, no. 2: 194–212. https://doi.org/10.1080/15021149.2020.1859348.

McTiernan, A., A. McCoy, J. Mendonca, H. Lydon, and S. Diffley. 2022. "The Implementation of Precision Teaching for the Improvement of Academic Skills: A Systematic Review of the Literature Over Thirty Years." *Behavioral Interventions* 37, no. 2: 505–528. https://doi.org/10.1002/bin.1852.

Mellard, D., M. McKnight, and J. Jordan. 2010. "RTI Tier Structures and Instructional Intensity." *Learning Disabilities Research & Practice* 25, no. 4: 217–225. https://doi.org/10.1111/j.1540-5826.2010.00319.x.

Morrell, M., R. Morrell, and R. M. Kubina Jr. 1995. "Using Precision Teaching to Enhance Direct Instruction Sight-Word Reading." *Journal of Precision Teaching and Celeration* 13, no. 1: 47–54.

Park, J., E. C. Bouck, and C. K. Josol. 2020. "Maintenance in Mathematics for Individuals With Intellectual Disability: A Systematic Review of Literature." *Research in Developmental Disabilities* 105: 103751. https://doi.org/10.1016/j.ridd.2020.103751.

Ross, D. E., J. Singer-Dudek, and R. Douglas Greer. 2005. "The Teacher Performance Rate and Accuracy Scale (TPRA): Training as Evaluation." *Education and Training in Developmental Disabilities* 40, no. 4: 411–423. https://doi.org/10.1177/215416470504000410.

Singer-Dudek, J., J. Speckman, and R. Nuzzolo. 2010. "A Comparative Analysis of the CABAS Model of Education at the Fred S. Keller School: A Twenty-Year Review." *Behavior Analyst Today* 11, no. 4: 253–265. https://doi.org/10.1037/h0100705.

Sparrow, S. S., D. V. Cicchetti, and D. A. Balla. 2005. *Vineland Adaptive Behavior Scales*. 2nd ed. AGS Publishing.

Stein, M., D. Kinder, K. Rolf, J. Silbert, and D. W. Carnine. 2018. *Direct Instruction Mathematics*. 5th ed. Pearson.

Sun, Y., and M. Stropoli. 2026. "Becoming Bookworms: Increasing Independent Engagement With Books Through a Classroom Reading Routine." *Tizard Learning Disability Review*: 1–9. https://doi.org/10.1108/TLDR-10-2025-0050.

Tarlow, K. R. 2017. "An Improved Rank Correlation Effect Size Statistic for Single-Case Designs: Baseline Corrected Tau." *Behavior Modification* 41, no. 4: 427–467. https://doi.org/10.1177/0145445516676750.

Verdun, V. R., D. M. Fienup, B. A. Chiasson, and R. D. Greer. 2022. "Arranging Peer-Tutoring Instruction to Promote Inference-Making." *Journal of Applied Behavior Analysis* 55, no. 2: 369–394. https://doi.org/10.1002/jaba.895.

Vostanis, A., C. Padden, and P. E. Langdon. 2022. "Investigating the Relationship Between Learning Channel Sets During the Mathematical Practice of Autistic Students." *British Journal of Special Education* 49, no. 3: 1467–8578. https://doi.org/10.1111/1467-8578.12408.

Vostanis, A., C. Padden, A. McTiernan, and P. E. Langdon. 2021. "Comparing the Minimum Celeration Line and the Beat Your Personal Best Goal-Setting Approaches During the Mathematical Practice of Students Diagnosed With Autism." *Journal of Behavioral Education* 32, no. 1: 21–50. https://doi.org/10.1007/s10864-021-09432-7.